A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Crystal Structure of VCA0593 from *Vibrio cholerae*, a putative guanylate dinucleotide ribonuclease

비브리오 콜레라균 이차전령물질의 분해효소 VCA0593의 구조적 특성 분석

February, 2018

Yongdae Jang

Department of Agricultural Biotechnology

College of Agriculture and Life Sciences

Seoul National University

석사학위논문

Crystal Structure of VCA0593 from *Vibrio cholerae*, a putative guanylate dinucleotide ribonuclease

지도교수 하 남 출

이 논문을 석사학위논문으로 제출함

2018년2월

서울대학교 대학원

농생명공학부

장 용 대

장용대의 석사학위논문을 인준함

2018년2월

위원장 강동현 (인)

부위원장 하남출 (인)

위원 최영진 (인)

Abstract

Vibrio cholerae is a serious foodborne pathogenic bacterium that infects estimated 2.8 million people each year. When invaded into the host, the bacteria cause severe diarrhea mainly by secreting cholera toxin. *V. cholerae* recognizes and adapts its local circumstance through regulating the cellular level of cellular c-di-GMP. In bacteria, c-di-GMP is degraded into the guanylate dinucleotide pGpG or two GMP molecules by the EAL domain containing proteins, and pGpG is degraded into two GMP molecules by oligoribonuclease or nanoRNase. In *Pseudomonas aeruginosa*, pGpG was accumulated when the activity of oligoribonuclease was repressed. The increased number of pGpG inhibited the break-down of c-di-GMP via a product inhibition mode. The nanoRNases consisting of DHH and DHHA1 domains have generally a broad substrate specificity, and thus they can hydrolyze c-di-NMPs as well as nanoRNAs. VCA0593 was identified in *V. cholerae* genome and characterized as the nanoRNase superfamily. Unlike other canonical nanoRNase superfamily, VCA0593 degraded only pGpG, but not the c-di-GMP and other nanoRNAs, according to the biochemical assay. To understand the molecular reason for the higher substrate specificity of VCA0593, I determined crystal structures of VCA0593 in this study. In

addition, I also determined the crystal structure of H282A mutant VCA0593 that lost its activity. His282 play an important role in binding to pGpG. Structural comparison of VCA0593 with the canonical nanoRNases and DNase containing the DHH and DHHA1 domains gave molecular sights into the higher substrate specificity of VCA0593. The unique dimeric structure provides a relatively rigid structure with a fixed active site cleft, which may be the molecular nature of the narrow substrate specificity.

This study confirmed the structural characteristics of VCA0593, which is expected to play an important role in regulation the c-di-GMP pathway of *V. cholerae*. Because inhibition of the break-down of c-di-GMP is related to the pathogenicity of *V. cholerae*, the development of the inhibitor based on the structure of VCA0593 may contributes the control of food poisoning induced by *V. cholerae*.

***Keywords*: *Vibrio cholerae*, c-di-GMP, pGpG, VCA0593, crystal structure**

***Student Number:* 2016-21742**

CONTENTS

LIST OF FIGURES

LIST OF TABLE

Ⅰ. Introduction

The fatal food-borne pathogen *Vibrio cholera* is curved rods shaped Gram negative bacterium that live in aquatic environment. The bacterium is a representative human pathogen which can cause severe diarrheal disease cholera [1]. *V. cholerae* infects estimated 2.8 million people each year and approximately 91,000 of them die in all over the world [1, 2]. *V. cholerae* causes disease in small intestine by secreting cholera toxin, an ADP-ribosylating toxin that induce serious watery diarrhea[3].

All bacteria, such as *V. cholerae*, utilize diverse signaling materials to recognize and adapt the circumstance. Both bis-(3′-5′)-cyclic dimeric adenosine monophosphate (c-di-AMP) and bis-(3′-5′)-cyclic dimeric guanosine monophosphate (c-di-GMP) play an important role for regulating the physiological state of cells as the second messenger transducing cell signaling [4-7]. Cellular level of c-di-AMP affects cell size, cell growth and sporulation. Cellular level of c-di-GMP affects motility, virulence, proteolysis, biofilm, and sessility of the bacteria. NanoRNAs typically designate very short RNA consisting of 2 - 5 nucleotides [8]. NanoRNAs are produced by degradation of mRNA. They are also utilized to decide the physiological state of the bacteria by serving as primers for transcription initiation and controlling gene expression [7-10]. Not only the mRNA but also c-di-GMP and c-di-AMP are decomposed into nanoRNAs. C-di-GMP and c-di-AMP are hydrolyzed into 5'-phosphguanylyl-(3',5')-guanosine (pGpG) and 5'-

phosphadenylyl-(3',5')-adenosine (pApA), respectively. Participated in regulation of the degradation pathway of c-di-GMP and c-di-AMP, pGpG and pApA have more important roles in regulation of the physiology of the bacteria than other nanoRNAs [11].

It is important to recognize and cope with the change of the surrounding environment to all living organisms. In particular, *V. cholerae* undergo the dramatic environmental change during transition to human hosts. To adapt to this environmental change, c-di-GMP is employed as a key secondary messenger in *V. cholerae*, which governs the physiological change between motile and sessile states [4]. Moreover, c-di-GMP are ubiquitously used as a secondary messenger in many other cellular signal pathways, involved in virulence, antibiotic resistance, and cell division in *V. cholerae* [12]. High cellular level of c-di-GMP induces high level synthesis of various adhesins and exopolysaccharides resulting in sessility and biofilm formation [4]. At a low level of c-di-GMP, high expression level of flagella was expressed together with their elevated activity, resulting in a high motility of the bacteria [4]. Furthermore, high level expression of acute virulence genes involved in secretion of the cholera toxin are induced when the cellular c-di-GMP level is low [2, 4]. Thus, the cellular level of c-di-GMP should be tightly controlled by the elaborate synthesis and degradation pathways for *V. cholerae* to adapt to changing environments, which is relevant to pathogenesis of the bacteria.

To achieve homeostasis of c-di-GMP, c-di-GMP is synthesized by dinucleonylate

cyclases and degraded by diverse phosphodiesterases that are structurally distinguished. The cyclases in synthesizing c-di-GMP commonly contain so-called GGDEF domain-containing enzymes by producing, one molecule of c-di-GMP from two GTP molecules [6, 7]. C-di-GMP is degraded into two GTP molecules by one or two step pathways. The one-step pathway usually employs phosphodiesterase or nanoRNases which have broad substrate specificity hydrolyzing both c-di-GMP and pGpG. At the two-step pathway employing two different enzymes, c-di-GMP is degraded to 5'-phosphoguanylyl-(3',5')-guanosine (pGpG) as an intermediate by a phosphodiesterase enzyme (called PDE-A), which typically contains EAL domain. Then pGpG is finally hydrolyzed to two molecules of GMP by another class of phosphodiesterase (called PDE-B) or oligoribonuclease or nanoRNase [12]. In *Pseudomonas aeruginosa*, an oligoribonuclease Orn played the primary role in degradation of pGpG, and the inhibition of Orn accumulated the level of pGpG. The excess pGpG substantially extended the half-life c-di-GMP by inhibiting the PDE-A that is required for the break-down of c-di-GMP through a product inhibition mechanism [6, 11].

VCA0593 was predicted or screened as a c-di-GMP and pGpG binding protein in *V. cholerae* genome by a genome-wide system approach [4, 5]. The amino acid sequence of VCA0593 suggested that it belongs to the nanoRNaseA (NrnA) superfamily. In recent study, VCA0593 was identified as a pGpG degrading ribonuclease by specific hydrolysis of pGpG (Young-Ha Park, Jae-Woo Lee, and

Yeong-Jae Seok, unpublished results). VCA0593 failed to hydrolyze other cyclic dinucleotide second messengers or other nanoRNAs except pGpG unlike canonic NrnA superfamily members [13-16].

The NrnA superfamily members commonly contain the DHH domain and DHHA1 domain. The DHH domain was named after the conserved motif of aspartate-aspartate-histidine residues, which are the catalytic residues for the phosphoesterase activity. The DHHA1 stands from the DHH domain associated domain 1, because the DHHA1 domain is usually paired with DHH domain in a single polypeptide. The DHHA1 domain contains a serial glycine-glycine-glycine-histidine residues, called the GGGH motif, which are related to the RNA or DNA binding function [14-17]. In addition, the NrnA superfamily members contain the partially conserved RxRxR motif, serial arginine-x-arginine-x-arginine residues related to the substrate binding, in the DHHA1 domain [15]. The canonical NrnA enzymes exhibit a broad specific activity which can hydrolyze nanoRNAs, such as pGpG, and pApA, and even some of them can hydrolyze c-di-AMP and c-di-GMP [17]. Five canonical NrnA structures have been determined to date. *Bacillus subtilis* NrnA (PDB ID: 5J21), *Bacteroides fragilis* nanoRNase (PDB ID 3W5W), *Mycobacterium smegmatis* NrnA (PDB ID: 4LS9), *Mycobacterium tuberculosis* Rv2837c c-di-NMP phosphodiesterase (PDB ID: 5CEU), and *Thermotoga maritima* phosphodiesterase (PDB ID: 5O1U) [13-15, 18].

To gain structural and functional information of VCA0593 which is distinguished

from canonical NrnA enzymes, I undertook to determine the crystal structure of VCA0593. In this study, I successfully purified the protein in the *E. coli* expression system, obtained the single crystals and finally determined the crystal structures of the VCA0593 protein. I discuss how VCA0593 exhibits the hydrolyzing activity only on pGpG by comparing with the canonical NrnA and the DHH-DHHA1 containing DNase structures. This study would contribute to the development of regulators for *V. cholerae*.

Ⅱ. Materials and Methods

2.1. Plasmid construction

All cells used in this study were provided by Drs. Young-Ha Park and Yeong-Jae Seok at Seoul National University. The recombinant plasmids (pETDuet –VCA0593 and pETDuet –VCA0593-H282A) were transformed into *Escherichia coli* strain BL21(DE3) using the heat-shock method and incubated overnight at 37°C for protein production.

2.3. Purification of VCA0593

2.3.1. Overexpression

The recombinant proteins used in this study were inoculated by picking a single colony from *E. coli* strain into 25ml of LB medium with 50μg/ml ampicillin (Duchefa, Netherlands) and incubated at 37°C overnight. The cells were cultured in 2.0 L of LB medium including appropriate antibiotics until an OD600 of 0.8 and protein production was induced with 0.5 mM IPTG at 30°C. Cells were harvested 7 ~ 8 h after induction, and the cell pellet was resuspended with 50 ml of lysis buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 2 mM 2-mercaptoethanol. The cell pellet was disrupted by Sonicator, and then the cell lysate was acquired by centrifugation at 13,000 rpm for 30 min.

2.3.2. Affinity chromatography

The cell lysate containing recombinant proteins was incubated with 2 ml of Ni^{2+}-NTA agarose resin and rolled in a column for 1 h at 4°C. The Ni^{2+}- NTA agarose resin was washed with lysis buffer containing 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20 mM Imidazole (pH 8.0), and 2 mM β-mercaptoethanol to purify the protein. Elution was carried out by using lysis buffer supplemented with 250 mM Imidazole (pH 8.0). The size and purity of the protein were measured by SDS-PAGE. (Fig. 1)

2.3.3. Size exclusion chromatography (SEC)

The eluted proteins were further purified by size exclusion chromatography (HiLoad Superdex 200 26/600; GE Healthcare). The column was calibrated using calibration kit containing several standard preparations (Fig. 2, GE Healthcare). The column was pre-equilibrated with buffer containing 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), and 2 mM β-mercaptoethanol. The fractions containing the VCA0593 protein were concentrated up to 9 mg/ml using a centrifugal filter concentration device (Millipore, Temecula, CA; 30 kDa cutoff). The purity of protein was estimated by SDS-PAGE. The proteins sample were stored frozen at -80°C.

2.4. Crystallization

Initial crystallization of VCA0593 was performed using automated crystal screening device MOSQUITO by sitting-drop vapour-diffusion method. Commercially available sparse-matrix screening solutions (MCSG, Anatrace) was

selected for the initial crystallization screening trials at 14°C. The protein solution (0.2 μl, 10 mg/ml) was mixed with a reservoir solution (0.2 μl) and equilibrated against 60 μl of the reservoir solution in a 96-well crystallization plate. Small cuboid crystals were obtained using the condition containing 0.2 M ammonium sulfate, Bis-Tris:HCl (pH 8.0), 25% (vol/vol) PEG 3350, 2 mM Tris(2-carboxyethyl) phosphine (TCEP). The crystallization conditions were further optimized using hanging-drop diffusion method under a reservoir solution containing 0.2 M ammonium sulfate, Bis-Tris:HCl (pH 8.5), 20% (vol/vol) PEG 3350, 10 mM manganese chloride, 2 mM TCEP at 14℃. In the final optimization experiments, 1 μl of the protein solution (10 mg/ml) was mixed with 1 μl of the reservoir solution, and equilibrated against 500 μl of the reservoir solution in a 15-well plate. Thick plate-shaped crystals for data collection appeared in 3-4 days (Fig. 3). The crystal of VCA0593-H282A was obtained in same condition with wild type VCA0593.

2.5. Structure determination and refinement

The crystals were transferred to 2 μl of cryo-protection buffer containing the reservoir solution and additional 30% (wt/vol) (±)-2-methyl-2,4-pentanediol (MPD) for 1 min, and then were flash-cooled in liquid nitrogen for data collection under cryogenic conditions. To solve the phasing problem, 10 mM zinc chloride was added in cryo-protection buffer of several crystals. The datasets of VCA0593 and VCA0593-H282A were collected on a direct X-ray detector Pilatus 6M (Fig. 4, Dectris, Switzerland), equipped in beamline 5C of Pohang Accelerator Laboratory,

Republic of Korea, at a wavelength of 0.9795 Å and 1.0082 Å. The datasets of Zn soaked crystal were collected at a wavelength of 1.2822 Å The crystal of VCA0593 belonged to the space group $P2_12_12$, with unit cell dimensions of a =68.4 Å, b = 148.3 Å, and c = 58.8 Å. The crystal of VCA0593-H282A belonged to the space group $P2_12_12$, with unit cell dimensions of a =68.4 Å, b = 148.9 Å, and c = 58.9 Å. The Zn soaked crystal of VCA0593 belonged to the space group $P2_12_12$, with unit cell dimensions of a =70.9 Å, b = 148.6 Å, and c = 58.0 Å. The program HKL-2000 was employed to process, merge, and scale the diffraction datasets [19]. Datasets of the crystals of VCA0593 were diffracted up to a resolution of 1.9 Å and datasets of the crystals of VCA0593-H282A were diffracted up to resolution 1.55 Å. Data collection statistics are provided in Table 1

Anomalous signals from four Zn sites were found in each subunit, and the resulting electron density map was sufficiently clear to build an initial model using the programs PHENIX and COOT [20, 21]. To solve the crystal structure of VCA0593 protein at 1.90 Å resolution, molecular replacement was carried out against the native diffraction dataset using the initial model as a search model in the program MOLREP in the CCP4 suite [22].

Ⅲ. Results

3.1. Structural determination of wild type VCA0593

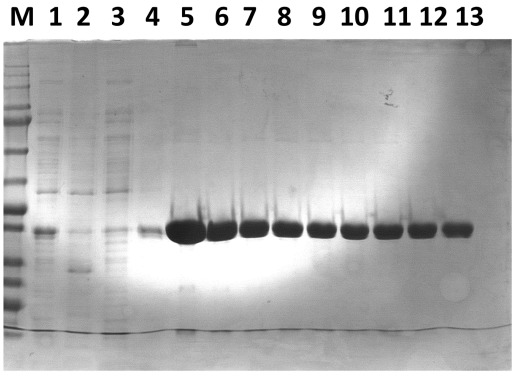
The wild type VCA0593 was successfully produced in the *E. coli* expression system using the expression vector pET-DUET, and purified using two subsequent chromatography steps consisting of the Ni^{2+}-NTA affinity chromatography, and gel filtration chromatography (Fig. 1 and 2).

The purity of the final sample was measured >98% judged by on the SDS-PAGE and the yield was 5 mg per 1.6 L culture. The protein band appeared in a molecular weight of ~36 kDa on the gel. The size exclusion chromatography indicated that VCA0593 is a dimeric protein in solution (Fig. 2).

The concentrated protein sample was subjected to the crystallization trials. The crystals of VCA0593 grew under several conditions from initial screening trials in an automated protein crystallization screening kit. The optimized crystals were obtained from 0.2 M ammonium sulfate, Bis-Tris:HCl (pH 8.5), 20% (vol/vol) PEG 3350, 10 mM manganese chloride, and 2 mM TCEP by hanging drop vapour diffusion method at 14°C. A highly brilliant undulator X-ray beam was utilized to collect high resolution dataset. The final dataset for the native protein exhibited 99.3% completeness at 1.9 Å resolution. The crystal lattice belonged to the primitive orthogonal space group, and analysis of the diffraction along the h, k and l axes

further revealed a space group of $P2_12_12$ with unit cell parameters of a = 68.9, b = 149.0, and c = 58.9 Å. (Table 1)

To obtain the phase information for the structural determination, Zn anomalous signals from Zn-soaked crystals was collected. The final dataset for the native protein exhibited 87.9% completeness at 2.3 Å resolution. The crystal lattice belonged to the primitive orthogonal space group, and analysis of the diffraction along the h, k and l axes further revealed a space group of $P2_12_12$ with unit cell parameters of a = 70.2, b = 148.6, and c = 58.0 Å. The resulting electron density maps allowed an almost complete model to be built, after which the structure was refined against the 1.9 Å-resolution dataset of the native protein. The asymmetric unit in the crystal contained two molecules.



Lane 1: Supernatant fraction
Lane 2: Precipitant fraction
Lane 3: Unbounded fraction
Lane 4: Washed fraction
Lane 5-13: Eluted faction

Figure 1. SDS-PAGE gel from Ni^{2+}-NTA chromatography.

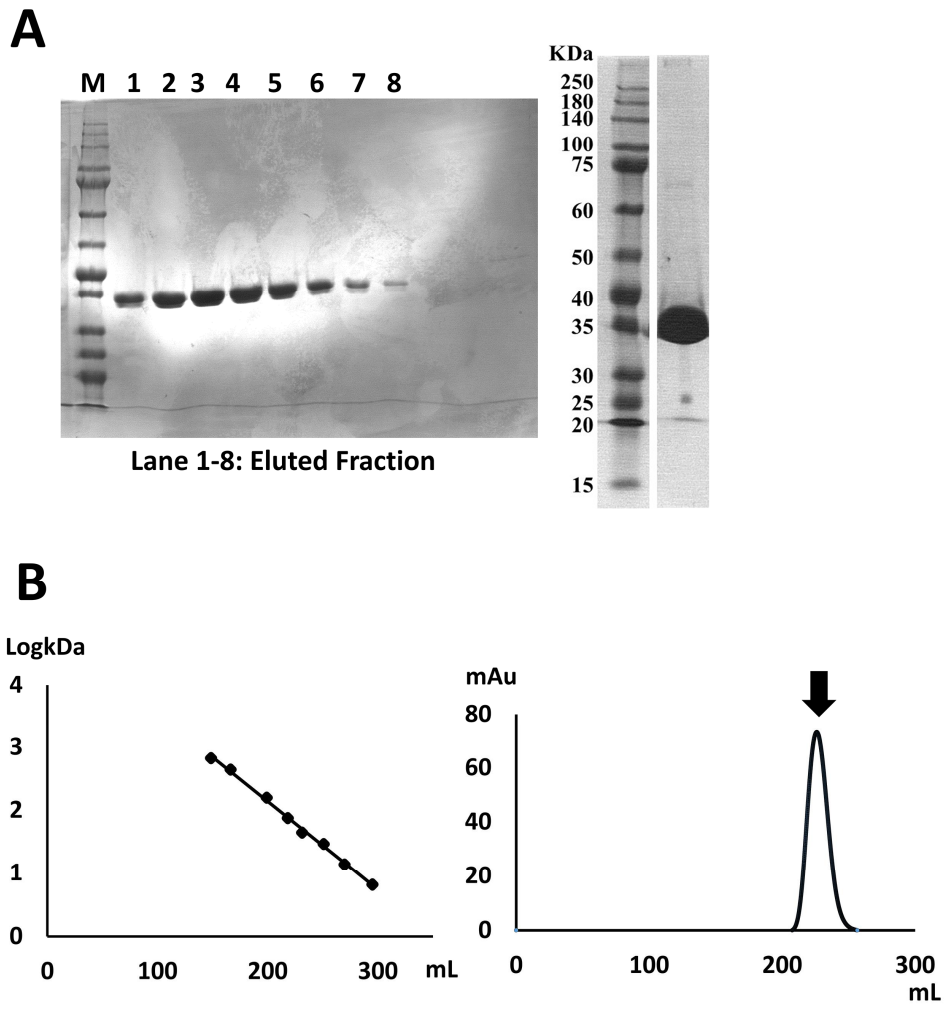


Figure 2. Purification result from SEC. A. SDS-PAGE gel from SEC (left) and final protein (right) **B.** standard curve of size exclusion chromatography (left); 148.5 ml (669 kDa) 166.21 ml: Ferritin (440 kDa) 199.1 ml: Aldolase (158 kDa) 218.23 ml: Conalbumin (75 kDa) 231.27 ml: Ovalbumin (44 kDa) 250.82 ml: carbonic anhydrase (29 kDa) 269.66 ml: Ribonuclease A (13.7 kDa) 294.97 ml: Aprotinin (6.5 kDa). Size exclusion chromatography elution

profile for measuring the molecular size of VCA0593 (right). The arrow at ~72 kDa indicates the molecular weight of the protein which corresponds to a homodimer (monomer size, ~36 kDa).

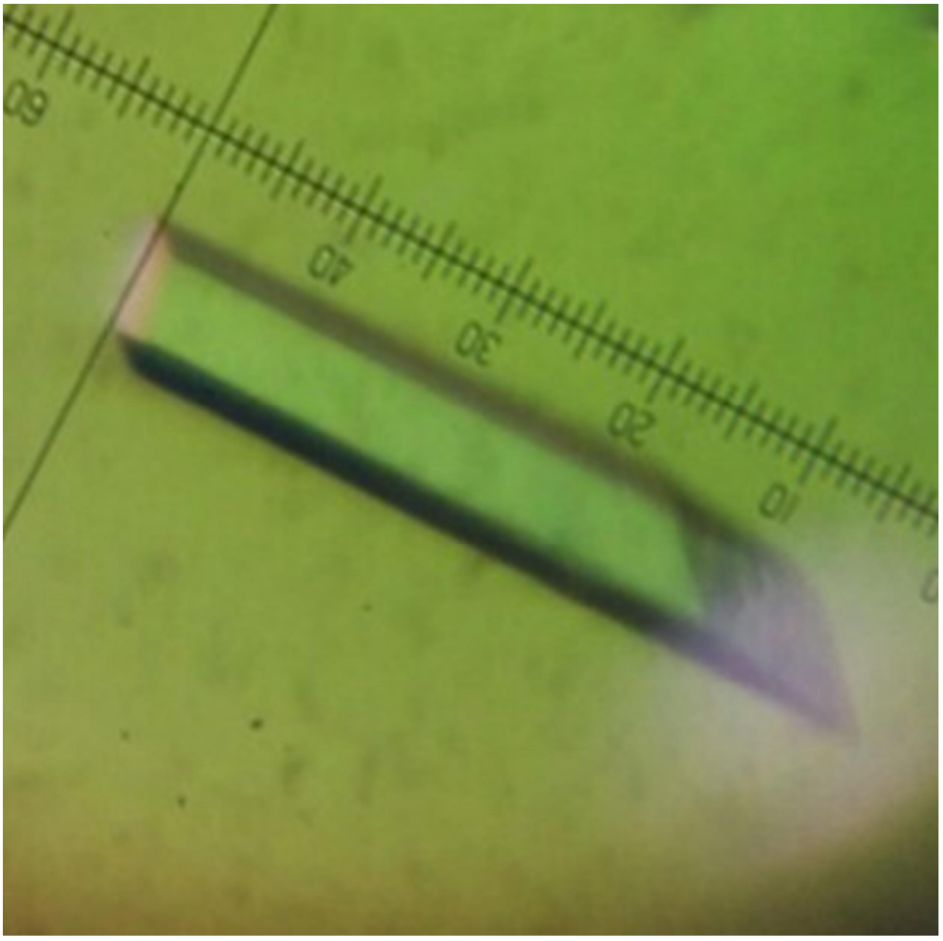


Figure 3. A crystal of VCA0593. The approximate dimension of the crystal is 0.35 x 0.35 x 0.05 mm.

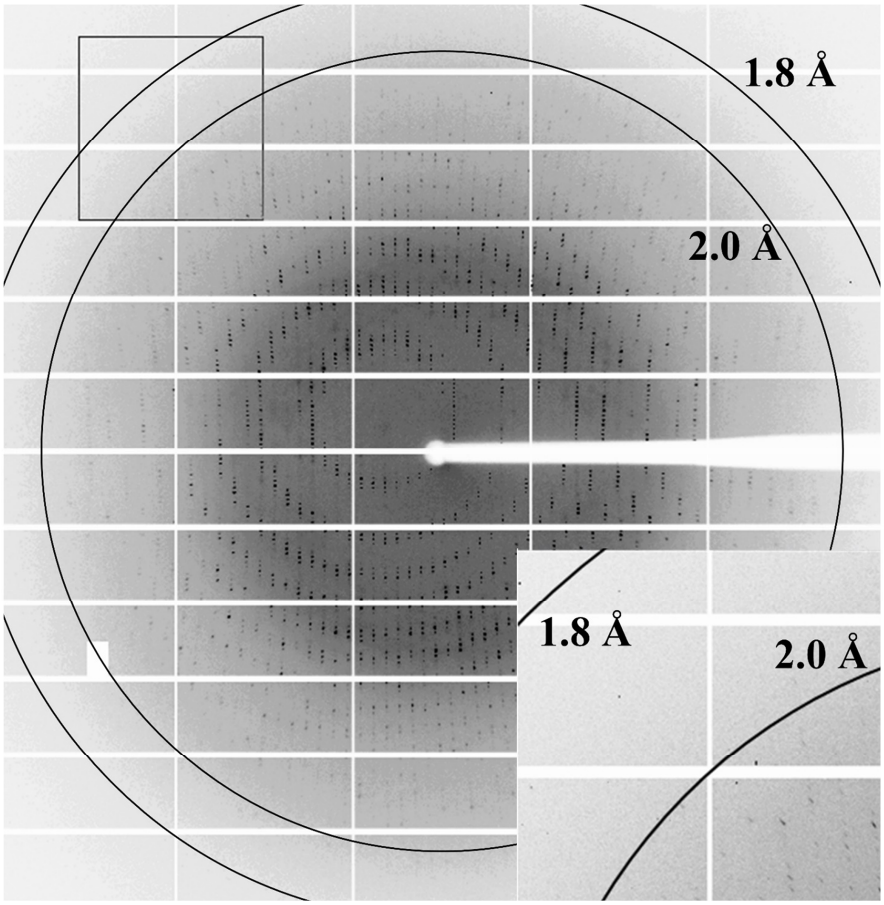


Figure 4. X-ray diffraction image of VCA0593.

	VCA0593	VCA0593-H282A	VCA0593_Zn
Data collection			
Beam line	PAL 5C	PAL 5C	PAL 5C
Wavelength (Å)	0.9795	1.0082	1.2822
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions			
a, b, c (Å)	68.4, 148.3, 58.8	68.4, 148.9, 58.9	70.9, 148.6, 58.0
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	38.25-1.90 (1.97-1.90)	38.29-1.55 (1.60-1.55)	50.00-2.30 (2.34-2.30)
R_{merge}	0.038 (0.206)	0.018 (0.155)	0.073 (0.013)
High resolution shell CC1/2	0.432	0.832	0.935
$I/\sigma I$	18.6 (16.9)	3.9 (2.5)	6.9 (4.9)
Completeness (%)	98.5 (94.7)	99.0 (94.5)	87.9 (82.2)
Redundancy	7.7 (4.2)	9.6 (6.3)	8.7 (5.8)
Refinement			
Resolution (Å)	1.90	1.55	
No. of reflections	43967	84228	
R_{work}/R_{free}	0.2058/0.2517	0.2059/0.22378	
No. of total atoms	4840	4964	
Wilson B-factor (Å)	20.50	11.56	
R.M.S. deviations			
Bond lengths (Å)	0.007	0.007	
Bond angles (°)	0.86	0.84	
Ramachandran plot			
Favored (%)	97.43	98.77	
Allowed (%)	2.57	1.23	
Outliers (%)	0	0	

Table 1. Statistics for X-ray data collection and refinement

* Values in parentheses are for the highest resolution shell.

** $R_{merge} = \Sigma_{hkl}\Sigma i|I_i(hkl) - [I(hkl)]|/\Sigma hkl\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and [I(*hkl*)] is the average intensity of the *i* observations.

*** R_{free} calculated for a random set of 10% of reflections not used in the refinement

3.2. The structure of the VCA0593 protomer

Like canonical NrnA proteins, the protomer of VCA0593 contains a N-terminal DHH domain (residues 1-181) and C-terminal DHHA1 domain (residues 206-310). Two domains are connected by linker α helix (α11, residues 182-205). The DHH domain is formed by ten α helices (α1, α2, α3, α4, α5, α6, α7, α8, α9 and α10) and five β strands (β1, β2, β3, β4, and β5). The DHHA1 domain is composed of three α helices (α12, α13 and α14) and six β strands (β6, β7, β8, β9, β10 and β11) (Fig. 5).

The overall structure of the VCA0593 protomer resembles the alphabet character "C". The DHH domain is in the upper part of "C", while the DHHA1 domain is in its lower part. The linking helix is in the middle of the "C". The central groove is found between the DHH and DHHA1 domains in VCA0593 protomer. The groove is decorated by the positively charged amino acid residues in its surface. This positively charged groove appears to be related to the substrate binding of VCA0593 (Fig. 5).

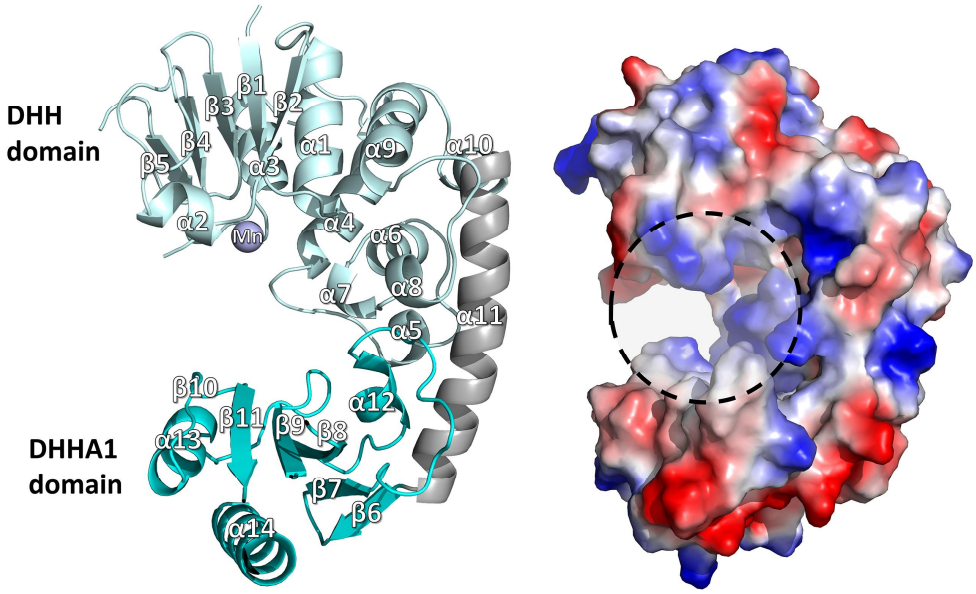


Figure 5. Structure of VCA0593 protomer. Ribbon representations of the monomeric structure of VCA0593 (left). The DHH domain is colored light cyan, the DHHA1 domain is colored cyan, and α11 is colored gray. Charged surface representation of the monomeric structure of VCA0593 is presented at right. Positive charge is colored blue, and negative charge is colored red. Mn^{2+} ion is presented as purple sphere.

3.3. The dimeric assembly of the VCA0593

The asymmetric units of VCA0593 contained two protomers that are closely interact, suggesting a homodimeric assembly. This dimeric assembly of VCA0593 is supported by the size exclusion chromatography result (Fig. 2 and 6). The two DHHA1 domains in the dimeric forms a tight interaction despite no direct interaction between DHH domains. Two sulfates are located in the groove of one protomer (Fig. 7).

In the structure of VCA0593, the dimerization of the DHHA1 domains forms the major dimeric interaction. Polar interactions are formed by Asn252, Thr287, and Gln289 on β11 of each protomer. A loop connecting β8 and β9 in the DHHA1 domain of one subunit is inserted between the DHH domain and DHHA1 domain of another subunit forming hydrophobic interactions. Met245, Val254 and Ala256 in one DHHA1 domain and Phe248 in the loop connecting β8 and β9 of another subunit were involved in the hydrophobic interaction.

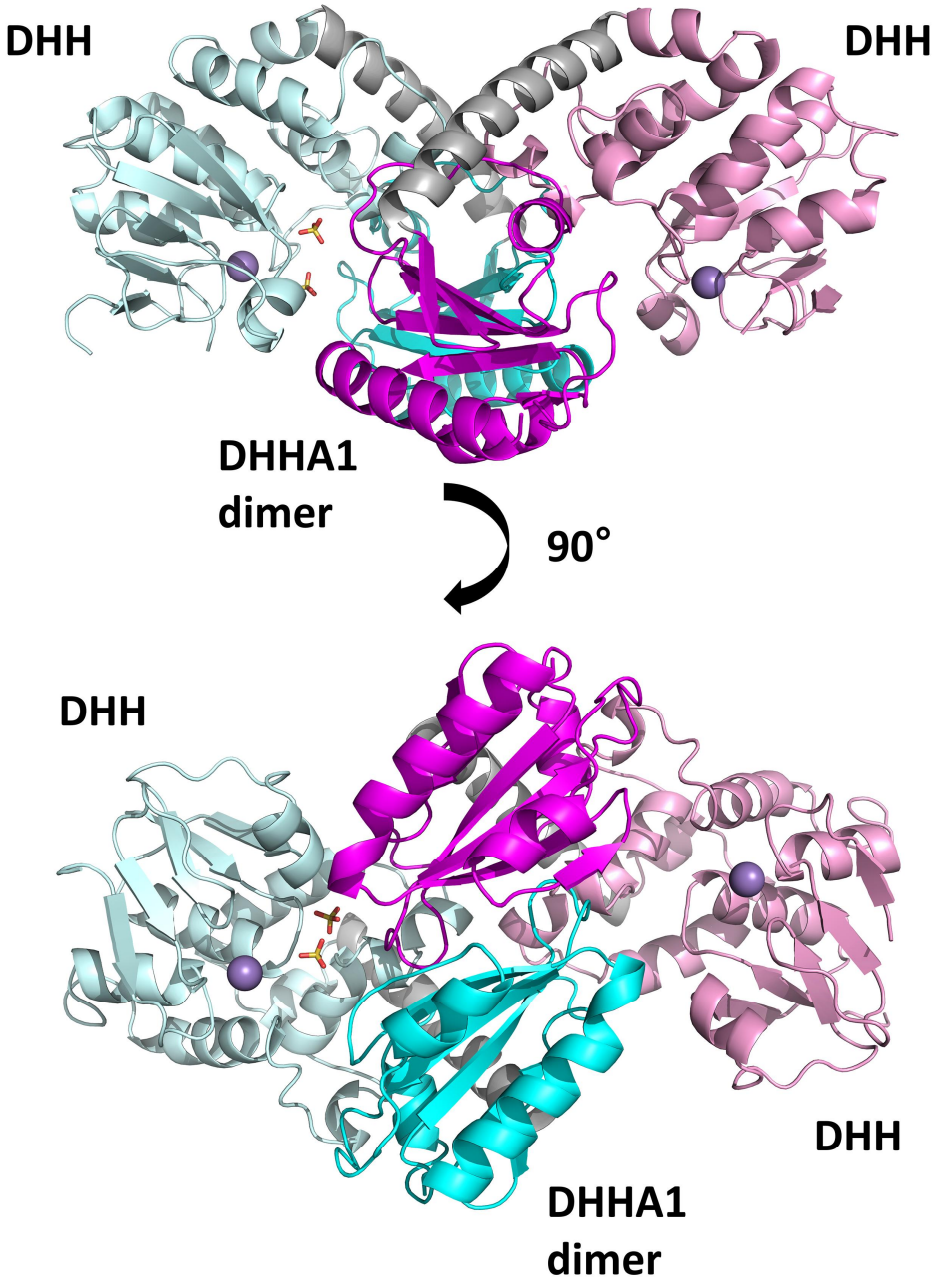


Figure 6. Structure of VCA0593 homodimer. Ribbon representations of the dimer structure of VCA0593. The two molecules are presented in

different colors as follows: Protomer A (DHH domain; light cyan, DHHA1 domain; cyan, α11 helix; grey), and protomer B (DHH domain; pink, DHHA1 domain; magenta, α11 helix; grey). Two sulfates are colored yellow. Mn^{2+} ion is presented as purple sphere.

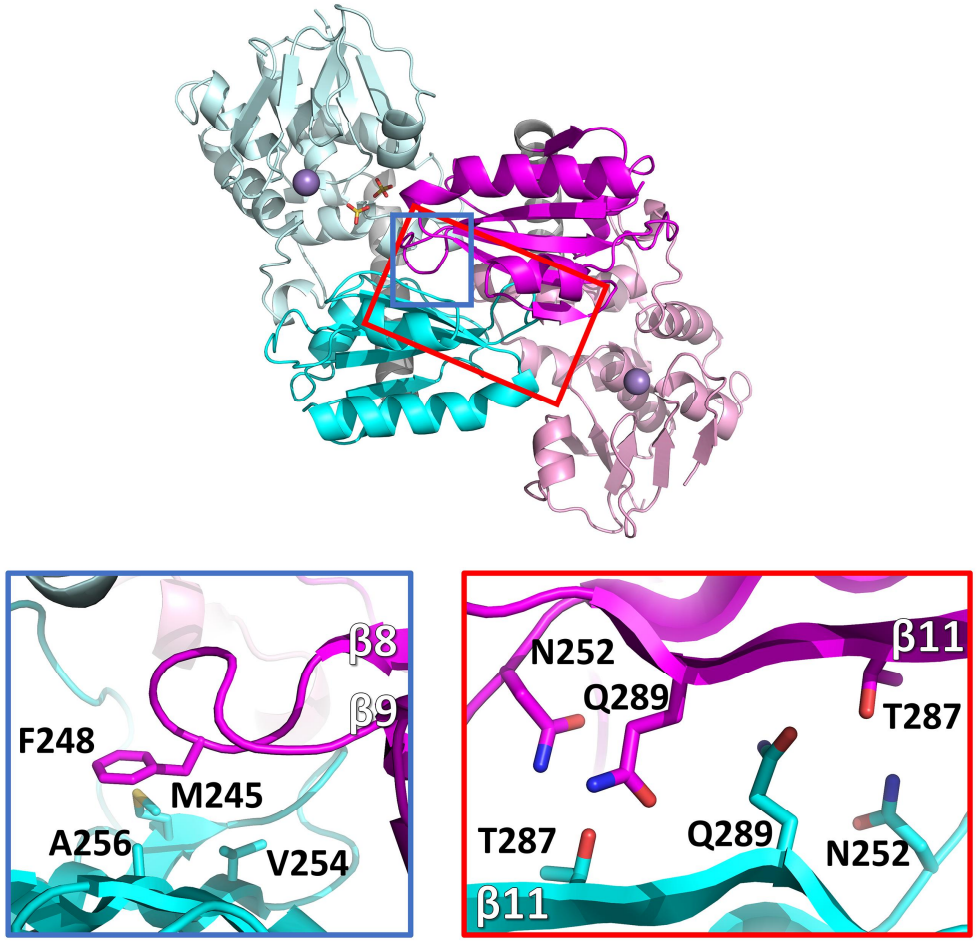


Figure 7. Dimeric interaction between two protomers. Major dimeric interaction is represented on the overview of VCA0593 dimeric structure as a blue and a red square. The two protomers flow the same color scheme as in Fig. 7. Hydrophobic interaction is magnified in a blue box. Polar interaction is magnified in a red box.

3.4. The positively charged groove lined with the metal binding site in the DHH domain and GGGH motif in the DHHA1 domain

The conserved DHH motif is in the positively charged groove region of the DHH domain, and are also exposed to the solvent in the dimeric unit of VCA0593. Like the DHH domains of the canonical NrnA proteins, a divalent metal ion is found in the structure in the DHH domain of VCA0593. The metal ion is coordinated by the two histidine residues (His68 and His69) in the DHH motif, Asp15, Asp121 and Asn55 within the DHH domain (Fig. 8). The coordination environment and the density map suggests that the metal ion is Mn^{2+}, which is well supported by the high concentration of $MnCl_2$ in the crystallization solution. Since the metal ion together with the adjacent residues are responsible for the phosphodiesterase activity in the DHH domain containing enzymes [14-16], this site would be also involved in the catalytic activity of VCA0593.

The conserved GGGH motif (residues 279-282) is in the DHHA1 domain of VCA0593. Especially, the histidine residue (His282) in the GGGH motif was known to play an important role in binding to single stranded RNA or DNA [16]. The GGGH motif also lines with the groove and exposed to the direction of the DHH motif. However, the RxRxR motif is not convered in the DHHA1 domain of VCA0593. Instead of RxRxR motif, Val254, Ala256, and Gly258 residues are located and these residues are involved in dimeric interaction. (Fig 8).

Two sulfate ions are found in the groove of one protomer (S1 and S2). S1 is exposed to the solvent and interacts with Arg11 and Arg 146, while S2 is embedded

in the DHH domain (Fig. 8). Because, the sulfate is structurally similar with the phosphate, the site of the sulfate can be the key to estimate the location of the phosphate in pGpG.

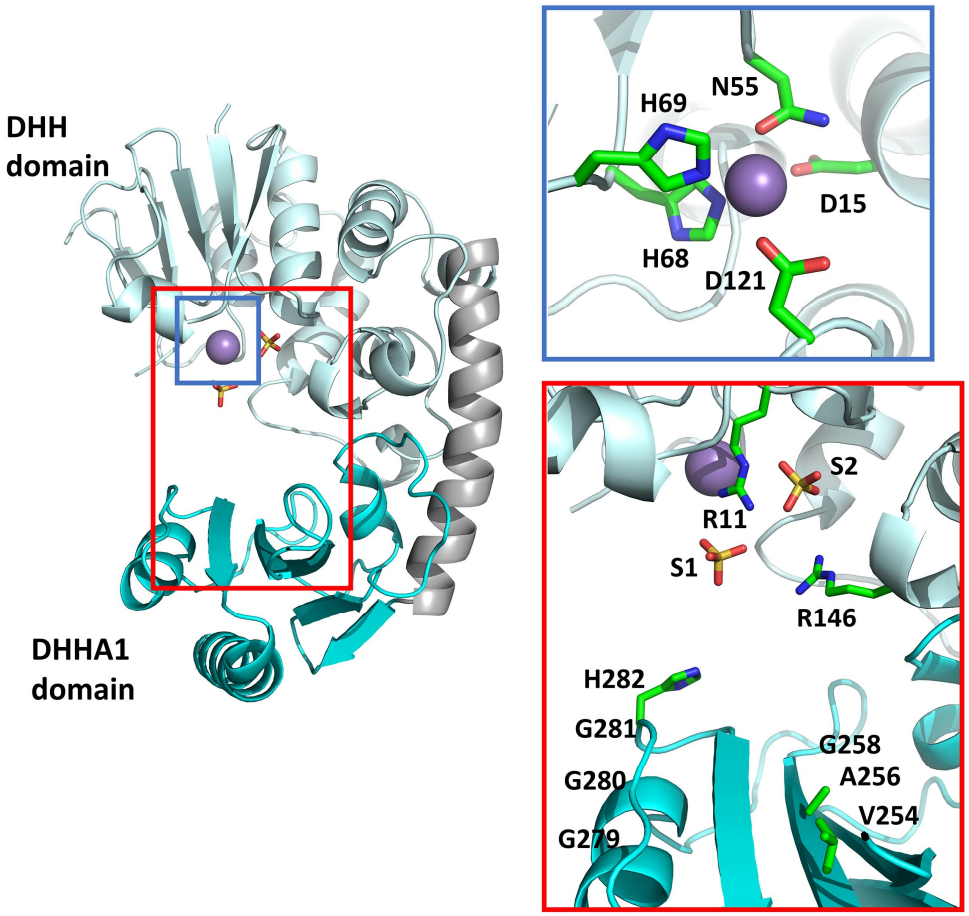


Figure 8. The positively charged groove and the metal binding site. The groove is presented in a red rectangle and the metal binding site is presented in a blue rectangle (left). Two sulfates are colored yellow. Mn^{2+} ion is presented as purple sphere. Detailed view of the groove and metal binding site is presented right. The metal binding site is presented in the blue box and the metal coordinating residues color green. Overall view of the groove is presented in the red box.

3.5. The role of His282 in VCA0593

According to an unpublished result from Young-Ha Park, Jae-Woo Lee, and Yeong-Jae Seok, the H282A mutant of VCA0593 completely lost its activity that hydrolyzes pGpG. To investigate the structural basis for the role of His282 in the GGGH motif of the DHHA1 domain, I determined the structure of H282A mutant using the crystals which grew under the same crystallization conditions as the wild type protein (Fig. 10). The crystal structure of H282A mutant was determined using molecular replacement by wild type VCA0593 as model structure. The structure of H282A mutant was refined the 1.55 Å-resolution dataset. To analysis the difference between the WT and the H282A mutant, the H282A mutant was superimposed onto an WT VCA0593. No evident difference was found both in the overall structure and active site except for the mutated residues (Fig. 10). The metal ion was also bound in the same region of the wild type protein. These results indicate that His282 only participates in binding with the substrate, but may not be involved in conformational change of the protein during the catalysis.

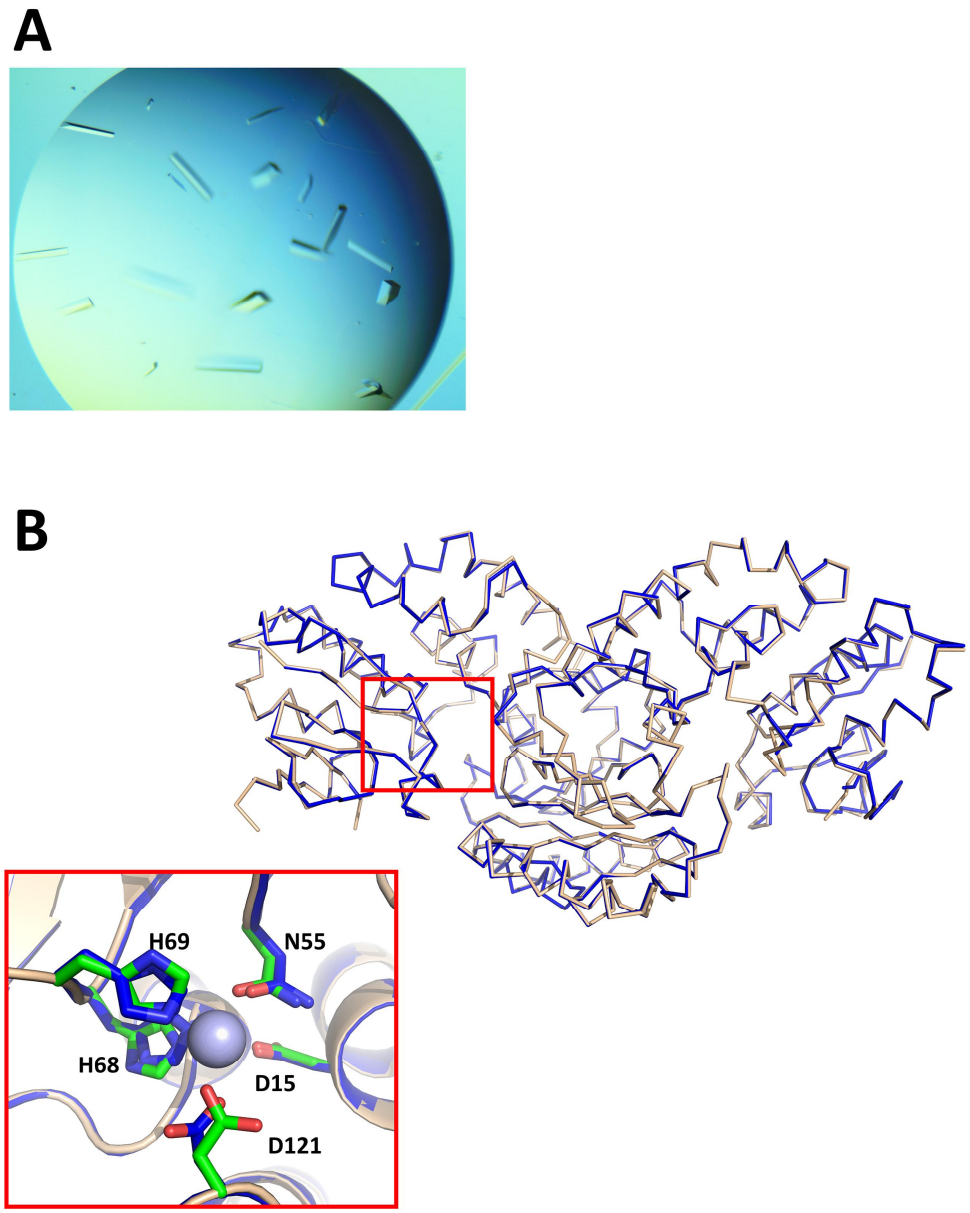


Figure 9. Crystals and overall structure of VCA0593-H282A. A. Crystals of VCA0593-H282A. **B.** Ribbon representations of the structural superposition of homodimer VCA0593-H282 (blue) onto wild type

VCA0593 (wheat). The metal binding site is highlighted as red rectangle, and the metal coordinating residues are presented in red box. Mn^{2+} ion is presented as purple sphere.

3.6. Structural comparison with a DNase exhibiting the 5' to 3' exonuclease specific to single-stranded DNA

I noted that the monomeric protein, *Thermus thermophiles* RecJ (ttRecJ), which exhibits a 5' to 3' exonuclease only acting on to single-stranded DNA. Like VCA0593, ttRecJ is composed of the DHH domain and the DHHA1 domain, which are connected by a long α helix [23]. Moreover, ttRecJ has the positively charged groove for binding to the substrate single-stranded DNA (Fig. 11). However, ttRecJ functions as a monomeric protein, which is different from the dimeric protein VCA0593 [23]. For ttRecJ to act on diverse length of single-stranded DNA, the size of the active site or the central groove should be adjustable depending on the size of the substrate DNA. In the monomeric protein ttRecJ, it is expected that the size of the groove could be changeable by a slight motion between the two domain. However, VCA0593 appears to allow only a very limited motion between the two domains because of the inter-subunit interaction in the dimer. Thus, this structural comparison suggests that the dimeric assembly of VCA0593 provides its narrower substrate specificity.

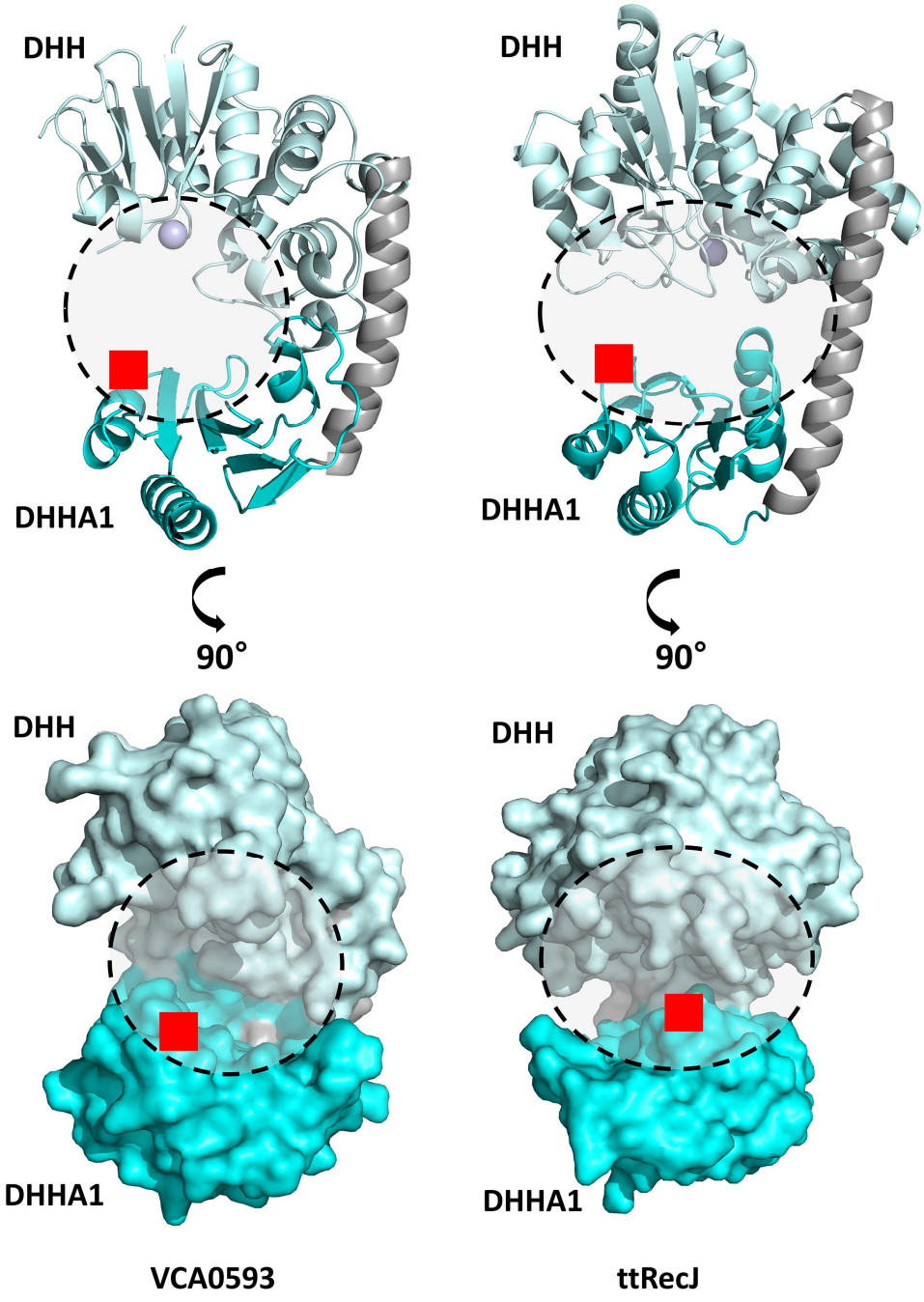

DHH
DHHA1
90°
DHH
DHHA1
90°
DHH
DHHA1
VCA0593
DHH
DHHA1
ttRecJ

Figure 10. Structural comparison of ttRecJ and VCA0593. Monomeric structure of ttRecJ and the VCA0593. The DHH domain is colored light cyan, and the DHHA1 domain is colored cyan in both structure. The groove between DHH domain and DHHA1 domain is presented as broken black circle. The location of Histidine of GGGH motif is presented as the red square.

3.7. Structural and functional comparison with the canonical NrnA protein

Given the broad substrate specificity of the canonical NrnA proteins, I analyzed the structures to gain insight into the molecular nature of the narrower substrate specificity of VCA0593. The dimeric structure of VCA0593 was compared with *Mycobacterium tuberculosis* Rv2837c (Rv2837c, PDB code: 5JJU), the representative canonical NrnA, in this study. Rv2837c has relatively broad substrate specificity which hydrolyze c-di-AMP, c-di-GMP, pGpG, and pApA [16]. VCA0593, and Rv2837c have the positively charged groove between the DHH and DHHA1 domain observed in each protomer, and the groove contains the histidine residue of the GGGH motif for the catalysis and the substrate binding.

In dimer structure of Rv2837c, two grooves in the dimer look opposite side. The dimeric interaction was composed by not only between two DHHA1 domains but two DHH domains in Rv2837c. However, two DHH domains of VCA0593 did not interact each other located in opposite site. In addition, two grooves in the dimer were looking same side and the groove of one protomer was partly screened by the inserted DHHA1 domain of another protomer in the dimeric structure of VCA0593 (Fig. 11). The inserted DHHA1 domains appear to limit the movement between the domains in the dimer. The DHH and DHHA1 domains are mutually fixed in the dimer, which might be related to the substrate specificity of VCA0593.

I next analyzed the temperature factors of VCA0593 by comparing with Rv2837c. In the high resolution structures, low temperature factors indicate high rigidity in the corresponding regions, and high temperature factor indicate the high structural

flexibility. The DHH and DHHA1 domain are connected by a flexible loop with high temperature factor in the structure of Rv2837c. However, the DHH and DHHA1 domains are connected by the long rigid α helix (α10) with lower temperature factors in the structure of VCA0593 (Fig 12). The difference of the connection of the DHH and DHHA1 domain suggests that two domains in Rv2837c interact each other with more flexibility than those of VCA0593. These structural analyses in terms of the temperature factors agrees well with the different domain arrangement in the dimers.

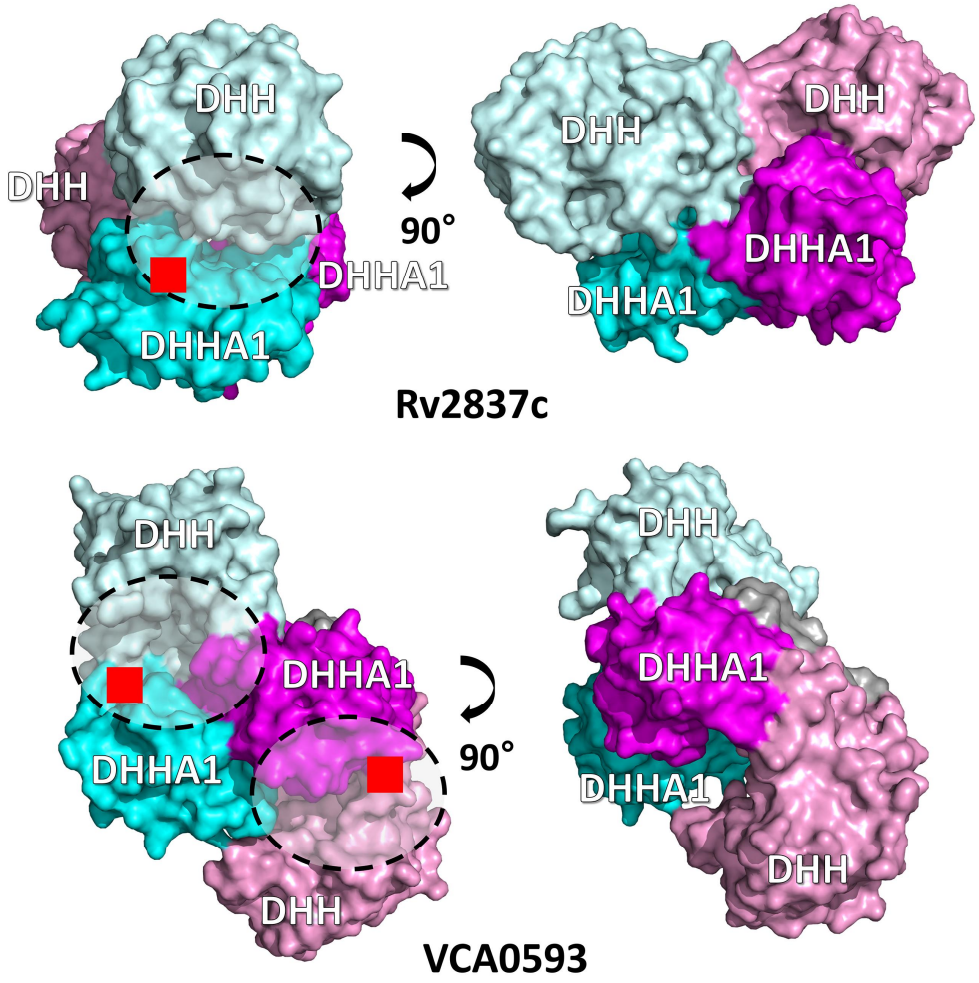


Figure 11. Comparison of dimeric structure between Rv2837c and VCA0593. Surface representations of homodimeric structures of VCA0593, and Rv2837c. The DHH domains are colored light cyan and pink. The DHHA1 domains are colored cyan and magenta. The groove between DHH domain and DHHA1 domain is presented as broken black circle. The location of Histidine of GGGH motif is presented as the red square.

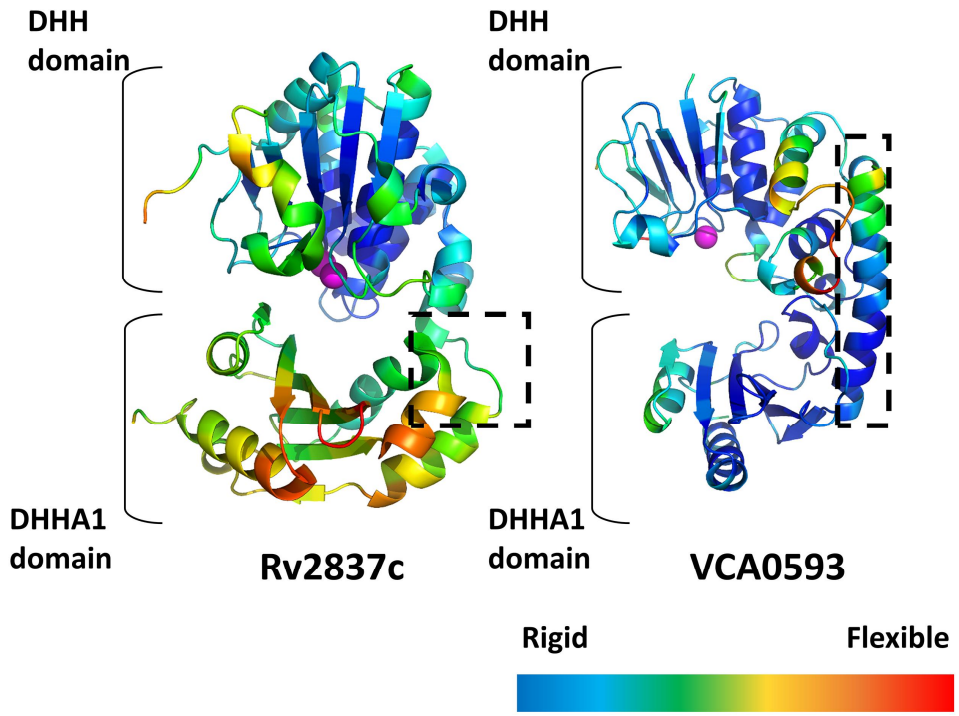


Figure 12. Comparison of temperature factor between Rv2837c and VCA0593. Ribbon representations of monomeric structures of VCA0593, and Rv2837c. Color of each structures represent the distribution of temperature factor and index bar is below the figure. Metal ions are represented as magenta spheres. The linker helix of VCA0593 and the linker loop of Rv2837c are highlighted as broken black rectangles.

Ⅳ. Discussion

This study presented the crystal structure of VCA0593 from *V. cholerae*, which is classified as the NrnA superfamily. VCA0593 is composed of the DHH domain containing the DHH motif and the DHHA1 domain containing the GGGH motif, as found in the canonical NrnA proteins. Two molecules are in one asymmetric unit forming a tighter dimer, which agrees well with the results of the size exclusion chromatography. Thus it could be concluded that this structure represents the functional oligomeric unit of VCA0593.

The crystal structure of a mutant VCA0593 harboring the substitution at His282 confirmed that the role of His282 is confined in the substrate binding although the mutation completely abolished the enzymatic activity of VCA0593.

To understand the narrower substrate specificity of VCA0593, I compared the structure of VCA0593 with the canonical NrnA protein and a DNase composed of the DHH domain and the DHHA1 domain. The monomeric DNase ttRecJ exhibited the similar structural compositions to the protomer structure of VCA0593. Since the substrate of ttRecJ is long single strand DNA, the substrate should penetrate the groove between the DHH domain and the DHHA1 domain [23]. Consistently, the two domains can be adjusted depending on the length of its substrates. However, the adjustment of the groove could not be expected because the two domains of VCA0593 is fixed within the dimer. Thus these findings indicate that the dimeric

formation is one of the important factor for the substrate specificity of VCA0593.

Then why do VCA0593 exhibit the narrow substrate specificity than the canonical NrnA proteins, although the canonical NrnA proteins are dimer? Interaction between the DHH and DHHA1 domain of VCA0593 is restricted by the DHHA1 domain inserted into the groove limiting the motion between two domains, while interaction of the DHH and DHHA1 domain of canonical NrnA free. Comparison of the temperature factor between VCA0593 and Rv2837c endorse the limited motion of VCA0593.

To suggest more detail mechanism of substrate binidng, the complex model of VCA0593 and pApA, the structural analogue of pGpG, was built using the complex structure of Rv2837c and pApA (Fig. 13A). In the complex structure of Rv2837c, the His312 of GGGH motif interact with the phosphate connecting two adenosines of pApA, and the Ser292 and Arg294 of RxRxR motif interact with the phosphate of 5' end like the complex structure of canonical NrnA and the substrate [13-16]. In the complex model of VCA0593, the inserted DHHA1 domain partly screened the groove and made it impossible to bind the substrate like other canonical NrnA (Fig. 13B). Because the sulfate has similar structure with the phosphate, it is inferred that the phosphate of pGpG can be located at the same site as sulfate. Arg11 and Arg146 interacting with the sulfate in the groove can act like the RxRxR motif. Two phosphate of pGpG may be located the sulfate and the histidine of GGGH motif, suggesting the different binding mechanism with other canonical NrnA (Fig. 13C). The different binding mechanism and structural rigidity may serve

as the narrow substrate specificity of VCA0593.

This study provided the structural evidence for the narrow substrate specificity of VCA0593. Because of the activity that specifically hydrolyze pGpG, VCA0593 may work as the key regulator of c-di-GMP signaling pathway in *V. cholerae*. Further studies need to know the detail mechanism how VCA0593 regulate the c-di-GMP signaling pathway. Because VCA0593 has the rigid structure produced by unique dimeric structure, and the linker helix, VCA0593 may be the ideal target for inhibitor study. The structural information in this study is helpful to develop the material that control the virulence of *V. cholerae* by regulating the c-di-GMP pathway.

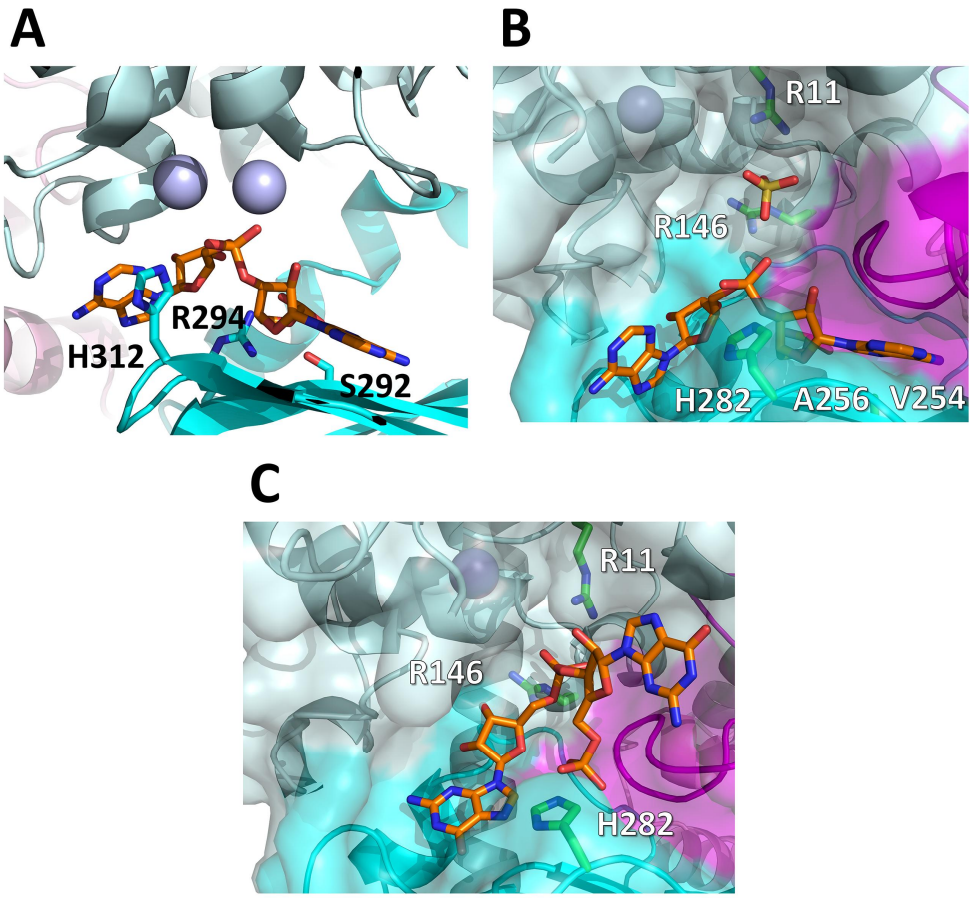


Figure 14. Comparison of the substrate binding of Rv2837c and VCA0593. A. Ribbon representation of the complex structure of Rv2837c and pApA. **B.** Ribbon and surface representation of the complex model of VCA0593 and pApA. **C.** Ribbon and surface representation of the complex model of VCA0593 and pGpG. The DHH domains are colored light cyan, the DHHA1 domains are colored cyan, and magenta.

Ⅴ. References

1. **Ali, M., et al.**, *The global burden of cholera.* Bull World Health Organ, 2012. **90**(3): p. 209-218A.
2. **Sloup, R.E., et al.**, *Cyclic Di-GMP and VpsR Induce the Expression of Type II Secretion in Vibrio cholerae.* J Bacteriol, 2017. **199**(19).
3. **Kaper, J.B., J.G. Morris, Jr., and M.M. Levine**, *Cholera.* Clin Microbiol Rev, 1995. **8**(1): p. 48-86.
4. **Hengge, R.**, *Principles of c-di-GMP signalling in bacteria.* Nat Rev Microbiol, 2009. **7**(4): p. 263-73.
5. **Conner, J.G., et al.**, *The ins and outs of cyclic di-GMP signaling in Vibrio cholerae.* Curr Opin Microbiol, 2017. **36**: p. 20-29.
6. **Jenal, U., A. Reinders, and C. Lori**, *Cyclic di-GMP: second messenger extraordinaire.* Nat Rev Microbiol, 2017. **15**(5): p. 271-284.
7. **Kalia, D., et al.**, *Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis.* Chem Soc Rev, 2013. **42**(1): p. 305-41.
8. **Nickels, B.E. and S.L. Dove**, *NanoRNAs: a class of small RNAs that can prime transcription initiation in bacteria.* J Mol Biol, 2011.

412(5): p. 772-81.

9. **Vvedenskaya, I.O., et al.**, *Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs.* Genes Dev, 2012. **26**(13): p. 1498-507.

10. **Goldman, S.R., et al.**, *NanoRNAs prime transcription initiation in vivo.* Mol Cell, 2011. **42**(6): p. 817-25.

11. **Cohen, D., et al.**, *Oligoribonuclease is a central feature of cyclic diguanylate signaling in Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2015. **112**(36): p. 11359-64.

12. **Corrigan, R.M., et al.**, *c-di-AMP is a new second messenger in Staphylococcus aureus with a role in controlling cell size and envelope stress.* PLoS Pathog, 2011. **7**(9): p. e1002217.

13. **Drexler, D.J., et al.**, *Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from Thermotoga maritima.* Structure, 2017.

14. **Uemura, Y., et al.**, *Crystal structure of the ligand-binding form of nanoRNase from Bacteroides fragilis, a member of the DHH/DHHA1 phosphoesterase family of proteins.* FEBS Lett, 2013. **587**(16): p. 2669-74.

15. **Schmier, B.J., C.M. Nelersa, and A. Malhotra**, *Structural Basis for*

the Bidirectional Activity of Bacillus nanoRNase NrnA. Sci Rep, 2017. **7**(1): p. 11085.

16. **He, Q., et al.**, *Structural and Biochemical Insight into the Mechanism of Rv2837c from Mycobacterium tuberculosis as a c-di-NMP Phosphodiesterase.* J Biol Chem, 2016. **291**(7): p. 3668-81.

17. **Mechold, U., et al.**, *YtqI from Bacillus subtilis has both oligoribonuclease and pAp-phosphatase activity.* Nucleic Acids Res, 2007. **35**(13): p. 4552-61.

18. **Valadares, N.F. and J. Woo**, *Mechanism of Rv2837c from Mycobacterium tuberculosis remains controversial.* J Biol Chem, 2017. **292**(32): p. 13480.

19. **Otwinowski, Z. and W. Minor**, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-26.

20. **Adams, P.D.**, et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

21. **Emsley, P. and K. Cowtan**, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

22. **Winn, M.D.**, *An overview of the CCP4 project in protein crystallography: an example of a collaborative project.* J Synchrotron Radiat, 2003. **10**(Pt 1): p. 23-5.

23. **Yamagata, A., et al.**, *The crystal structure of exonuclease RecJ bound to Mn2+ ion suggests how its characteristic motifs are involved in exonuclease activity.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5908-12.

Ⅵ. 국문 초록

비브리오 콜레라균은 해양환경에 서식하는 식중독 균으로 매년 전 세계적으로 약 280 만명이 감염되고 91000 명 정도가 목숨을 잃는다. 비브리오 콜레라균은 숙주 내로 침입하면 콜레라 독소를 이용해 심각한 설사를 유발하는 질병인 콜레라를 일으킨다. 비브리오 콜레라균은 세포 내의 c-di-GMP 농도를 통해 주변환경의 변화를 인식하고 적응하며, 세포 내의 c-di-GMP 농도는 합성과 분해에 의해 조절된다. 박테리아에서 C-di-GMP 는 EAL 도메인에 의해 pGpG 또는 2 개의 GMP 로 분해되고, pGpG 는 다시 올리고라이보뉴클레이즈 혹은 나노알엔에이즈에 의해 2 개의 GMP 로 분해된다. 녹농균에서 올리고라이보뉴클레이즈를 넉아웃하면 pGpG 가 축적되고 c-di-GMP 의 분해를 경쟁적으로 저해해 최종적으로 녹농균의 독성이 억제 된다는 것이 확인된 바 있다. 비브리오 콜레라균의 VCA0593 은 나노알엔에이즈에 속하는 효소로 pGpG 만을 특이적으로 분해한다고 알려졌다. 이전까지 연구된 나노알엔에이즈는 c-di-AMP, c-di-GMP, pGpG, pApA 등 상대적으로 넓은 기질특이성을 가지는 것으로 알려져 있었다. VCA0593 이 분자적 수준에서 pGpG 만을 분해하는 기질특이성을 가지는 이유와 역할에 대해서는 밝혀지지 않았다.

VCA0593 의 활성을 분자적 수준에서 연구하기 위해 X 선 결정학을 이용해 구조를 1.9 Å 해상도로 규명했다. 다른 연구집단에서 282 번 히스티딘 잔기를 알라닌으로 치환한 VCA0593 이 활성을 잃는다는 것을 확인해 그 원인을 분자적 수준에서 확인하고자 구조를 1.5 Å 해상도로 규명했고, 282 번 히스티딘이 기질과의 결합에 중요한 역할을 할 것이라고 결론 내릴 수 있었다. VCA0593 이 pGpG 만을 분해하는 기질특이성을 가지는 이유를 분석하고자 구조가 규명된 다른 나노알엔에이즈 및 DHH-DHHA1 도메인을 가지는 디엔에이즈와 구조를 비교해보았다. 그 결과, VCA0593 의 독특한 다이머 구조와 경직된 구조가 단백질의 움직임을 제한해 특이성에 영향을 줄 것으로 예측되었다.

본 연구를 통해 비브리오 콜레라균의 c-di-GMP 조절에 중요한 역할을 할 것으로 예상되는 VCA0593 의 구조적 특성을 확인했다. C-di-GMP 의 분해를 방해하면 비브리오 콜레라균의 독성을 제어할 수 있기 때문에, 구조를 기반으로 VCA0593 의 저해제를 개발한다면 비브리오 콜레라에 의한 식중독을 제어하는데 기여할 수 있을 것이다.

주요어: 비브리오 콜레라, c-di-GMP, pGpG, VCA0593, 결정구조